



HAL
open science

Mining tortured abbreviations from the scientific literature

Alexandre Clause, Guillaume Cabanac, Pascal Cuxac, Cyril Labbé

► **To cite this version:**

Alexandre Clause, Guillaume Cabanac, Pascal Cuxac, Cyril Labbé. Mining tortured abbreviations from the scientific literature. 2023. hal-04311600v2

HAL Id: hal-04311600

<https://hal.science/hal-04311600v2>

Preprint submitted on 24 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

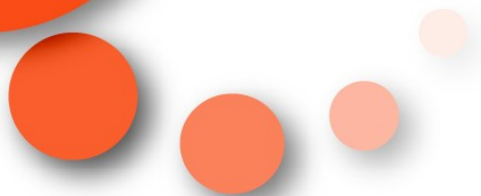
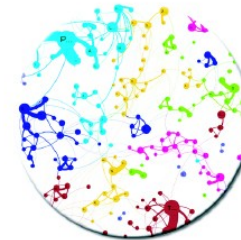


Distributed under a Creative Commons Attribution 4.0 International License



CNRS - Toulouse INP - UT3 - UT Capitole - UT2

Institut de Recherche en Informatique de Toulouse



Mining tortured abbreviations from the scientific literature

Alexandre Clause, Guillaume Cabanac, Pascal Cuxac, Cyril Labbé



UNIVERSITÉ
TOULOUSE III
PAUL SABATIER
Université
de Toulouse



Inist



LABORATOIRE D'INFORMATIQUE DE GRENOBLE



Université
Grenoble Alpes



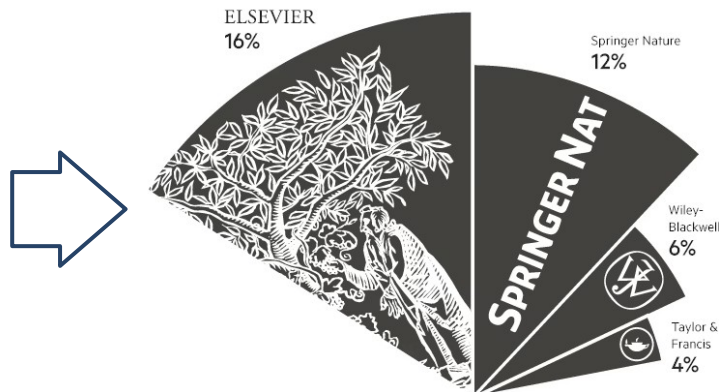
European Research Council
Established by the European Commission



Publish or perish



Elsevier top of the league
Global market share of journal articles by leading publishers, 2013 (%)



Source: company

<https://www.ft.com/content/93138f3e-87d6-11e5-90de-f44762bf9896>

voice recognition

deep neural network



Convolutional Neural Network. Artificial Neural Networks with numerous layers are termed as Deep Neural Networks or Deep Learning. It has been explored as one another key resource in recent years and has become quite well recognized in the literary community because of its efficiency to manage with huge amounts of data [17]. The most well-known **profound neural network** is the Convolutional Neural Networks (CNNs), which takes its name from operation of mathematical dimension from the matrixes termed convolution. Convolutional Neural Network (CNN) has various types of layers; it includes pooling, nonlinearity, and convolutional and fully connected layers. Convolutional Neural Network has pivotal outcomes over previous decades in an assortment of fields identified with **design**, **acknowledgment** from picture handling to **voice acknowledgment** [8]. The significant part of CNN is to get theoretical highlights when information proliferates towards the more **profound layers**. For instance, in picture characterization, the edge may be distinguished in the principal layers, and afterward the less difficult shapes in the subsequent layers, and afterward the **more elevated level** **high-level features**.

pattern recognition

deep layers

high-level features

pubpeer.com/publications/99826B46D2BA62C9FA87CAF11AE0FD

Tortured phrases

spectral local linearization method



All Images Videos Shopping Forums More

Scholarly articles for spectral local linearization method

A new **spectral local linearization method** for nonlinear ... - Motsa - Cited by 173

... conditions using **spectral local linearization method** - Sithole - Cited by 23

Spectral methods to solve nonlinear problems: A ... - Rai - Cited by 27

The spectral local linearization method. The unearthly neighborhood linearization technique (SLLM) can be used to unravel an arrangement of the non-comparable conditions. Within this method the criteria for logarithmic criteria are linearized using thoughts similar to the Gauss–Seidel approach.



ScienceDirect.com

<https://www.sciencedirect.com/science/article/pii>

Spectral methods to solve nonlinear problems: A review

About featured snippets • Feedback

<https://web.archive.org/web/20240521113315/https://www.google.com/search?client=firefox-b-d&q=spectral+local+linearization+method#ip=1>

(snapshot taken on May 21, 2024)

Tortured phrases

spectral local linearization method



All Images Videos Shopping Forums More

Scholarly articles for spectral local linearization method

A new **spectral local linearization method** for nonlinear ... - Motsa - Cited by 173

... conditions using **spectral local linearization method** - Sithole - Cited by 23

Spectral methods to solve nonlinear problems: A ... - Rai - Cited by 27



The spectral local linearization method. The **unearthly neighborhood linearization technique (SLLM)** can be used to unravel an arrangement of the non-comparable conditions. Within this method the criteria for logarithmic criteria are linearized using thoughts similar to the Gauss–Seidel approach.



ScienceDirect.com

<https://www.sciencedirect.com/science/article/pii>

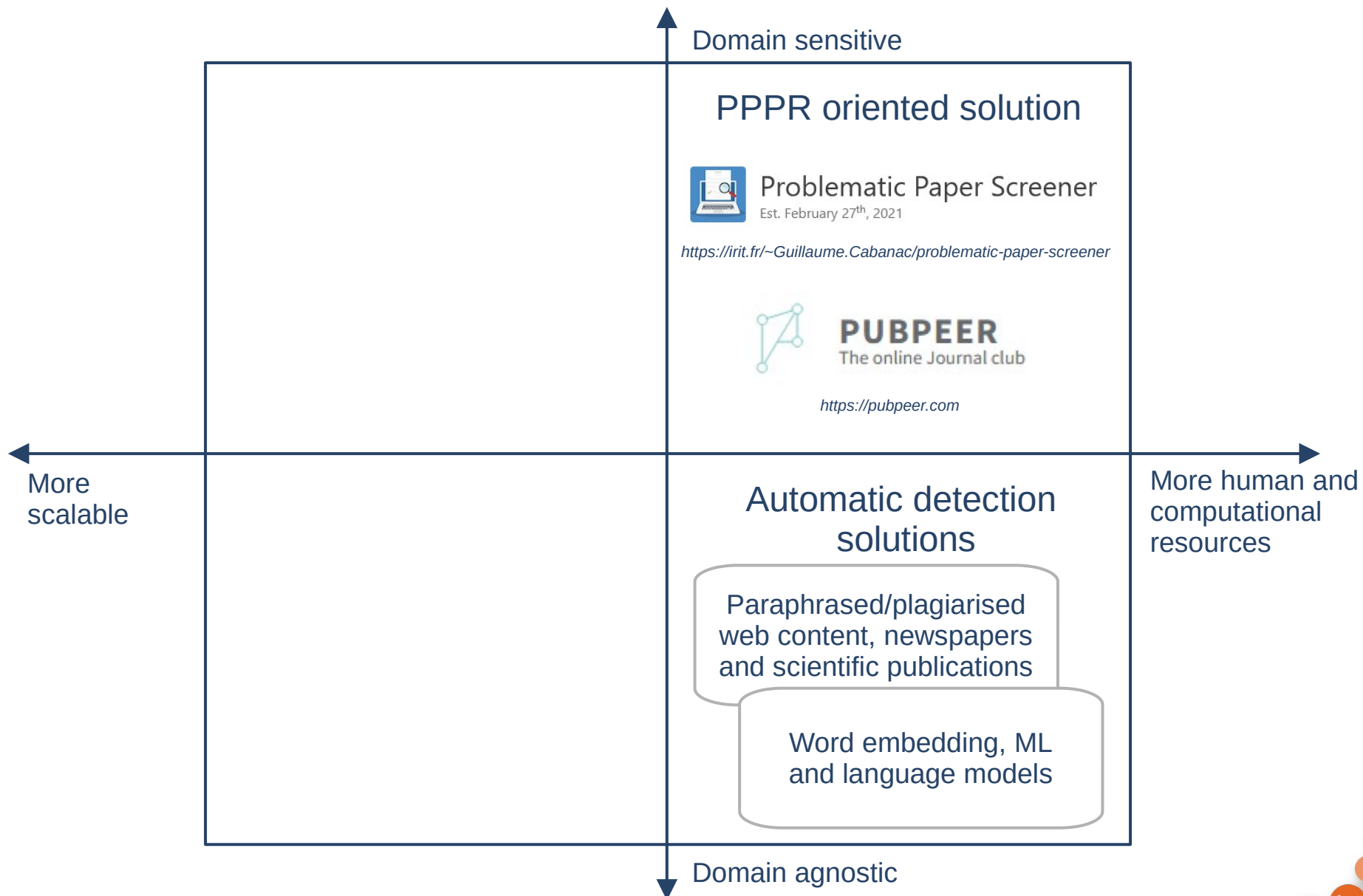
Spectral methods to solve nonlinear problems: A review

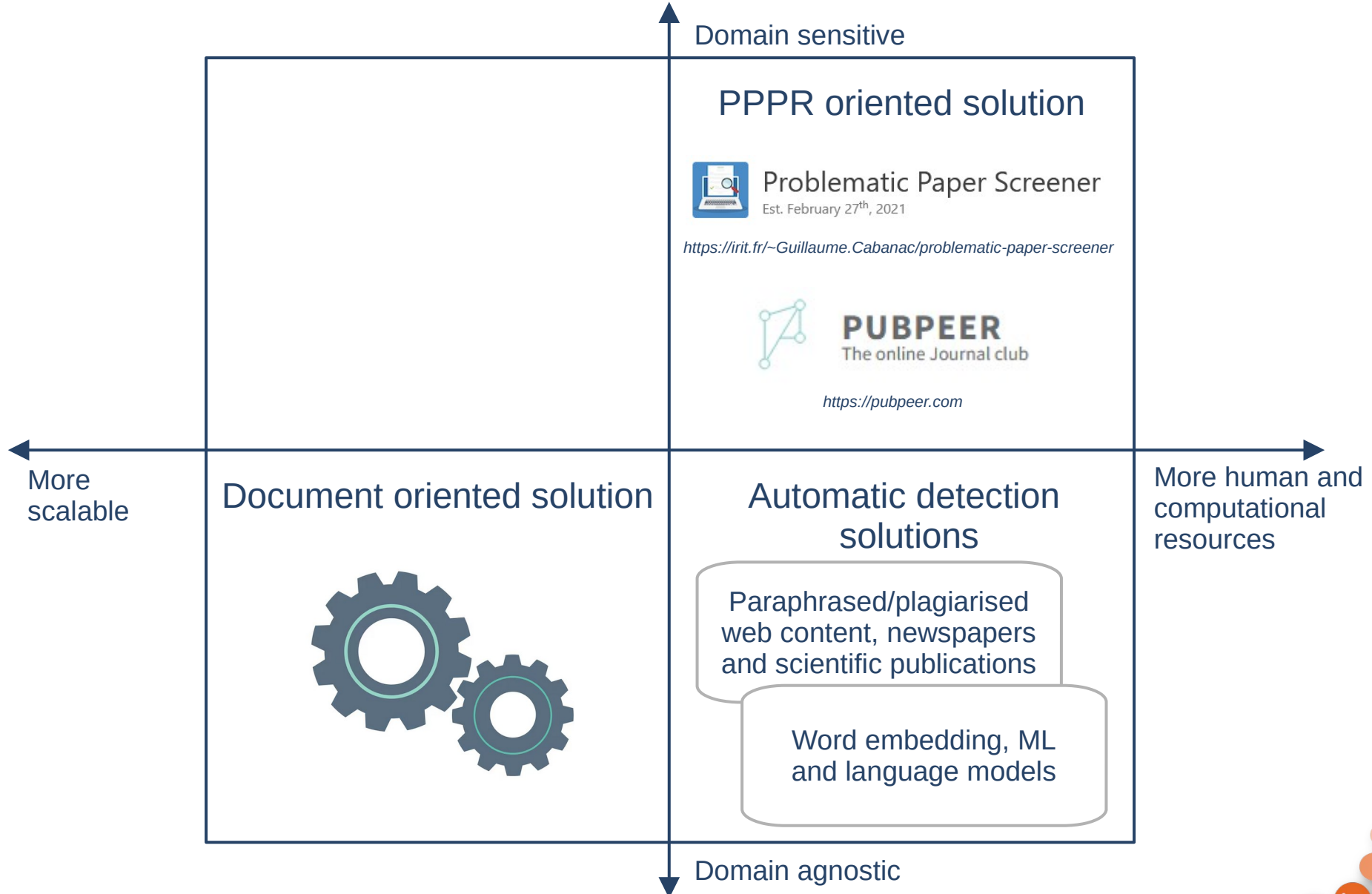
About featured snippets Feedback

<https://web.archive.org/web/20240521113315/https://www.google.com/search?client=firefox-b-d&q=spectral+local+linearization+method#ip=1>

(snapshot taken on May 21, 2024)

Existing solutions



Existing solutions

Proposed solution



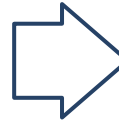
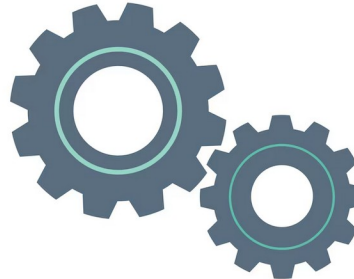
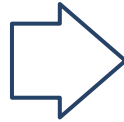
Draft



Preprint



Article



Submission (authors)

Editorial workflow (publishers)

Decision making (editors)

on IARPA's CS2 and NIST's IJB-A in both check and recognizable proof (for example search) and [9].

A half breed framework is introduced in which a convolutional neural organization and a logistic Regression Classifier are consolidated. Discriminant investigation is more effective when the ordinariness suspicious are fulfilled. The thoughts are finished on Yale face information base shows improved arrangement rates in more modest measure of time [10]. Class Activated Mapping (CAM) is utilized to show the areas that our organization uses to segregate between [11].

Dense SIFT and customary SIFT are considered in [12] and contrasted when combined and CNN highlights. Besides, an aggregator of the models is created. Results show the prevalence of CNN with Dense SIFT over ordinary CNN and CNN with SIFT. The exactness even expanded when all the models are collected which produces condition of confidence results on 73.4% and 99.1% on FER-2013 and CK- respectively.

We propose a grouping technique, called the **closest component line (NFL)** for **face acknowledgment**. Any two component purposes of individual face images by the element line (FL) going through the two focuses. The infers FL can catch big number of faces of the pictures than the first focuses and in this manner extends the limit of the accessible information base. The characterization depends on the closest separation from the question include highlight every FL. With a consolidated face information base, the NFL mistake rate is about 43.7-63.4% of that of the standard eigenface technique. Besides, the NFL accomplishes the most reduced mistake rate and answered to date for the ORL face information base [13].

The current works fail to announce the CNN designs that function admirably for **face acknowledgment** instead of examine the explanation. In this network, we address this issue by presenting CNN-based **face acknowledgment framework** (CNN-FRS) on a shared belief to make our work effectively reproducible. The three principal structures of CNN designs were utilizing LFW information. This paper quantitatively thinks about the designs of CNNs and assesses the impact of various usage decisions. We recognize a few helpful hypotheses of CNN-FRS. For example, the dimensionality of the **scholarly highlights** can be altogether decreased without unfriendly impact on **face acknowledgment exactness**. Likewise, a conventional measurement learning technique showing CNN-learned highlights is assessed. Analyses show two critical components to great CNN-FRS execution are the combination of numerous CNNs and metric learning [14].

A convolution neural organization (CNN) with **consideration instrument** that can see the impediment districts of the face and spotlight on the most discriminative unblocked locales. Thinking about various ReLU, we present two realizations of ACNN. aACNN just focuses on neighborhood facial patches. aACNN incorporates nearby portrayals at fit level with worldwide portrayal aspect level. Trial results show that ACNNs improve the acknowledgment precision on both the non-impediment and blocked countenance. Perception results show that, contrasted and the CNN without Gate Unit, ACNNs are fit for moving the consideration from the blocked patches to other related yet unimpeded ones. ACNNs additionally outperform other 'best-of-breed' techniques on face and video in the labourer appearance datasets under the cross-dataset assessment convention [15].

With the quickly advancement of the face acknowledgment (FR), warm **face acknowledgment** has gotten expanding consideration. In any case, the conventional techniques for warm **face acknowledgment** for the most part focus on the hand-made component plan, which requires more endeavour. **Face acknowledgment** is a **multimodal** task which requires

has moderately lower **acknowledgment rate**. A convolutional neural organization (CNN) design for warm **face acknowledgment** is presented. CNN is another kind of neural organization strategy which can consequently take in viable highlights from the **image information**. Investigation results on FER-2013 face information base show that our proposed CNN engineering accomplishes higher **acknowledgment rate** contrasted and the customary acknowledgment, for example, LBP, HOG and minutiae invariant [16].

Video-based **face acknowledgment framework** joins **invariant neural organization** (RNN) and 3D convolutional networks (3D) in a late-combination style. RNN and 3D encode appearance and movement data in an unexpected way. In particular, RNN takes appearance highlights removed by convolutional neural organization (CNN) over individual video outlines as info and encodes movement later, while 3D models appearance and movement of video at the same time. Joined with a sound module, our framework accomplished an **acknowledgment exactness** of 59.02% without utilizing any extra **face** marked video cuts preparing set, contrasted with 53.8% of the state-of-the-art [17].

Brood examinations show that consolidating RNN and 3D together can improve video-based **face acknowledgment** perceptibility [17].

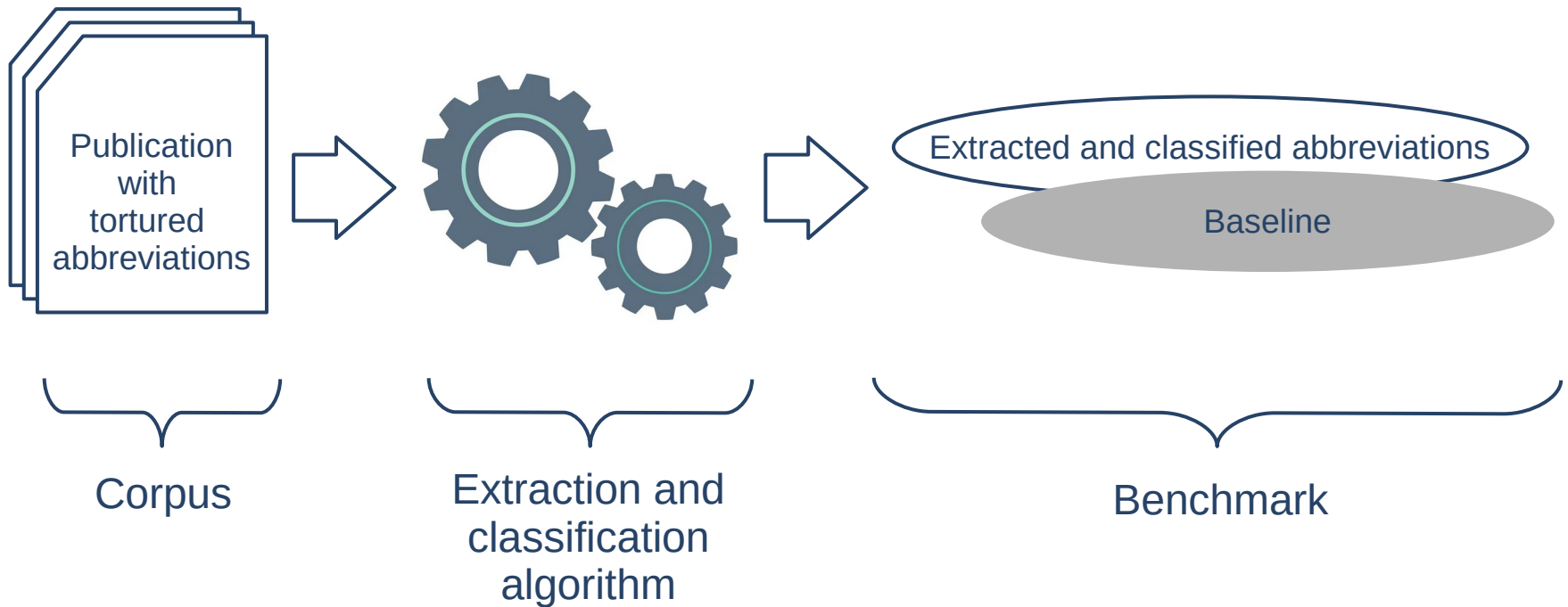
Perceiving human appearances is quite possibly the most well-known issues in the field of computer acknowledgment. Numerous methodologies and strategies have been tried and applied on the theme, particularly **neural organizations**. This paper proposed another method that can be supplemented at the lower part of a **neural organization** engineering as far as **face acknowledgment**, called **object-to-object** layer (OTL). To make more sure forecasts and arrangements, this **invariance layer** helps the **profound learning model** to indicate further discernible groups between various individuals (classes) by setting additional requirements on pictures of **face** similar individuals (inter-individual) while separating edges, contours, and features in individual between individuals. This proposed **object-to-object** layer improved the **acknowledgment exactness** on countenances by 2% [19].

Face acknowledgment is critical to certifiable applications, for example, video observation, human machine connection and security frameworks. When contrasted with conventional AI (*draw, net*), **profound learning** based techniques have indicated better exhibitions regarding precision and speed of preparing in **picture acknowledgment**. This paper proposes an altered Convolutional Neural Network (CNN) design by adding two standardization activities to two of the layers. The standardization activity which is bunch standardization gave increasing speed of the organization. CNN design was utilized to extract particular face highlights and Softmax classifier was utilized to characters faces in the **completely associated layer** of CNN. In the trial part, Georgia Tech Database indicated that the proposed approach has improved the **face acknowledgment exactness** with better acknowledgment results [21].

Face pictures showing up in sight and sound applications such as secure payment systems through cloud [21] [22], informal organizations and computerized management ordinarily display emotional posture, brightening, and appearance varieties, bringing about significant execution debasement for conventional **face acknowledgment** calculations. This paper proposes a complete **profound learning** structure to get the **face** and **posture** variations in the data. The proposed **profound learning** structure is made out of a bunch of extravagantly planned convolutional neural organizations (CNNs) and a three-layer stacked auto-encoder (SAE). The arrangement of CNNs separate covariate **face** highlights from **multimodal** information. At that point, these extracted highlights are connected to frame a high-dimensional component vector, whose measurements are compacted by SAE. The entropy of the CNNs are prepared utilizing a subset of 2000 subject, from the **face** and **posture** CASIA-WebFace information base which comprises of 1000000

<https://pubpeer.com/publications/2432C53B542C6B29AF449894368D61>





Bolster Vector Machine (SVM) → *Support* Vector Machine (SVM)

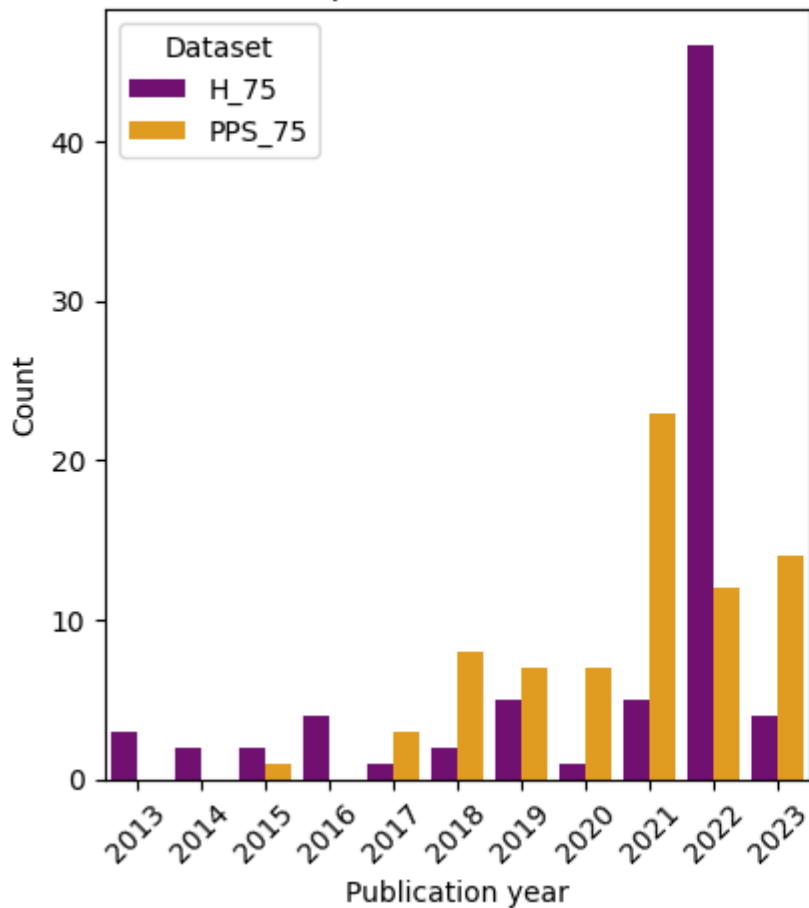
- **150** (PPS_75, dev + H_75, eval) open access **publications**
- **Published** between **2013** and **2023**
- Several **multidisciplinary** publications

Dataset → Class	PPS_75	H_75		
		H_G25	H_R25	H_T25
Genuine	608	345	179	124
Invalid	56	16	12	7
Tortured	480	17	96	87

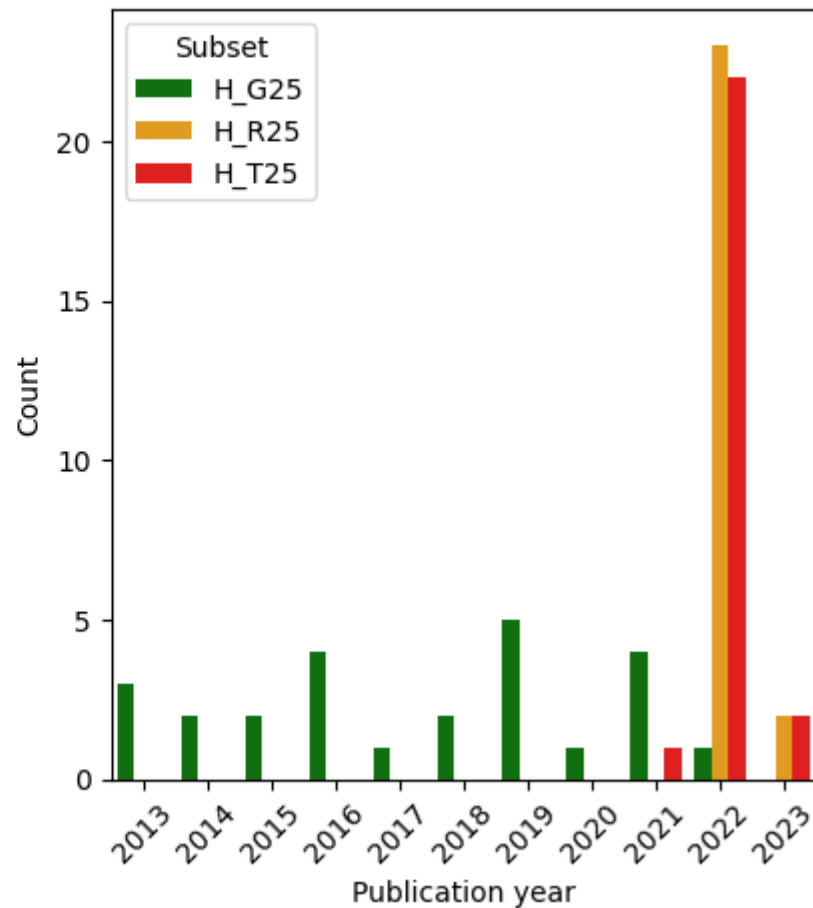
<http://doi.org/10.5281/zenodo.10492230>

Publication year distributions

Between development and evaluation datasets

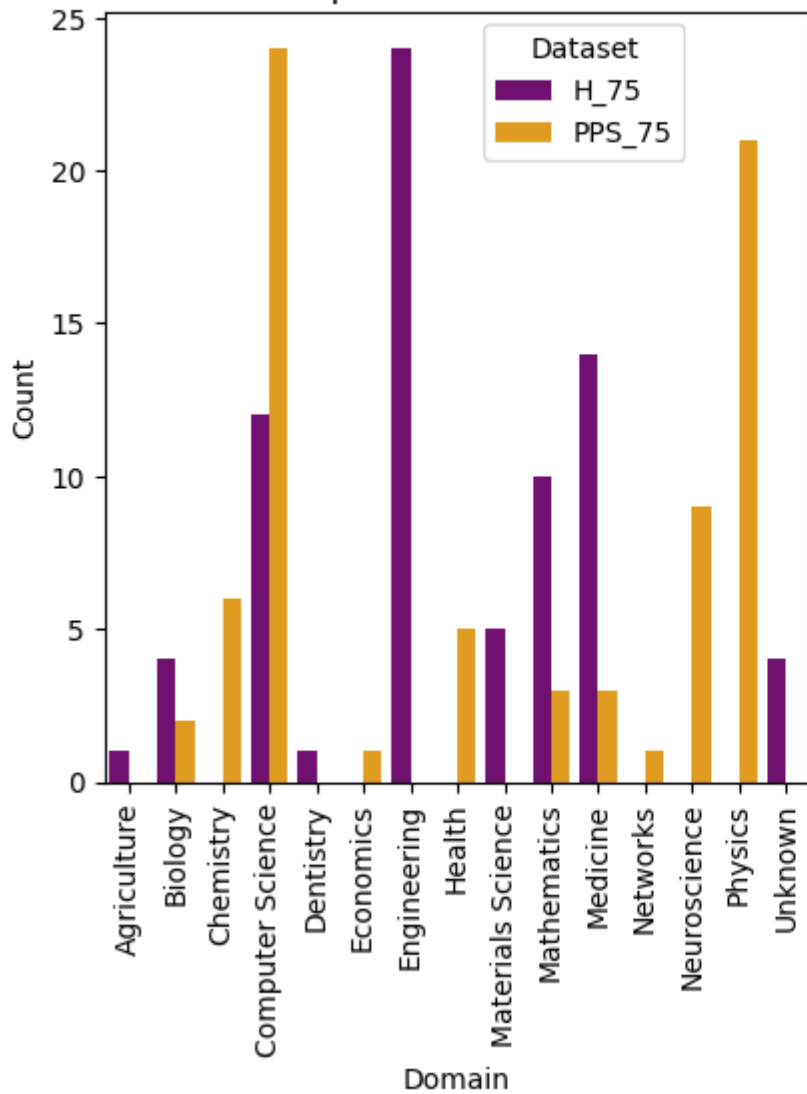


Between evaluation subsets

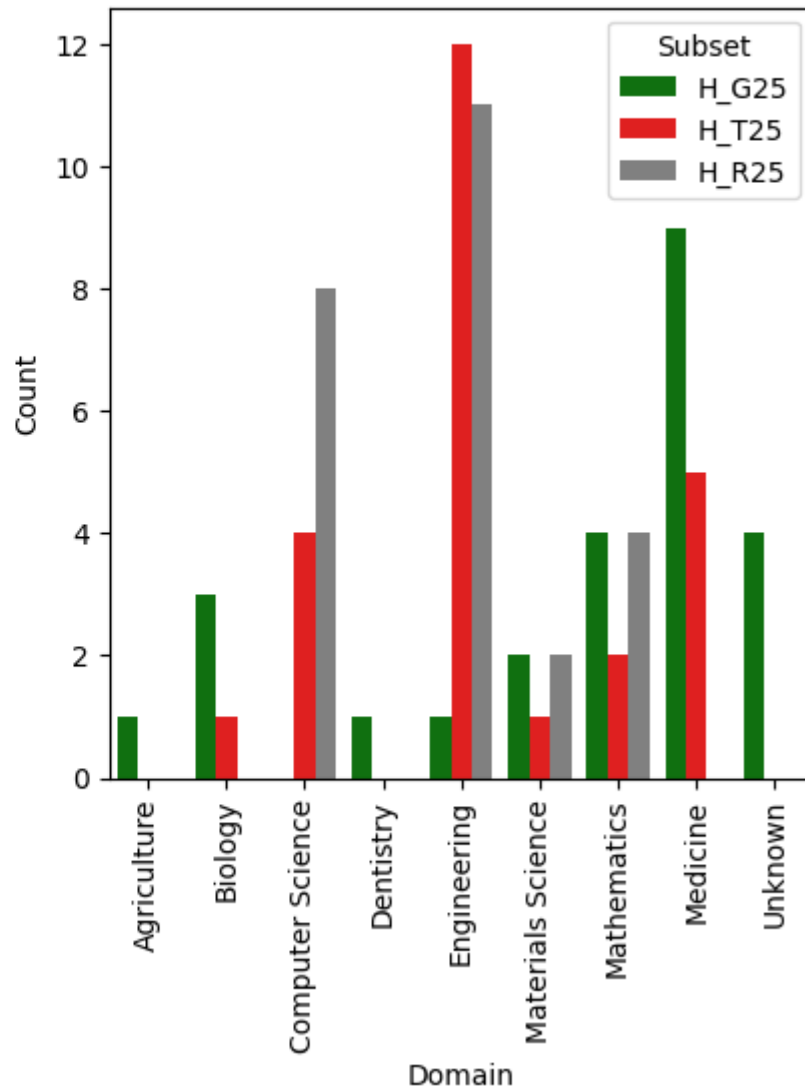


Domains distributions

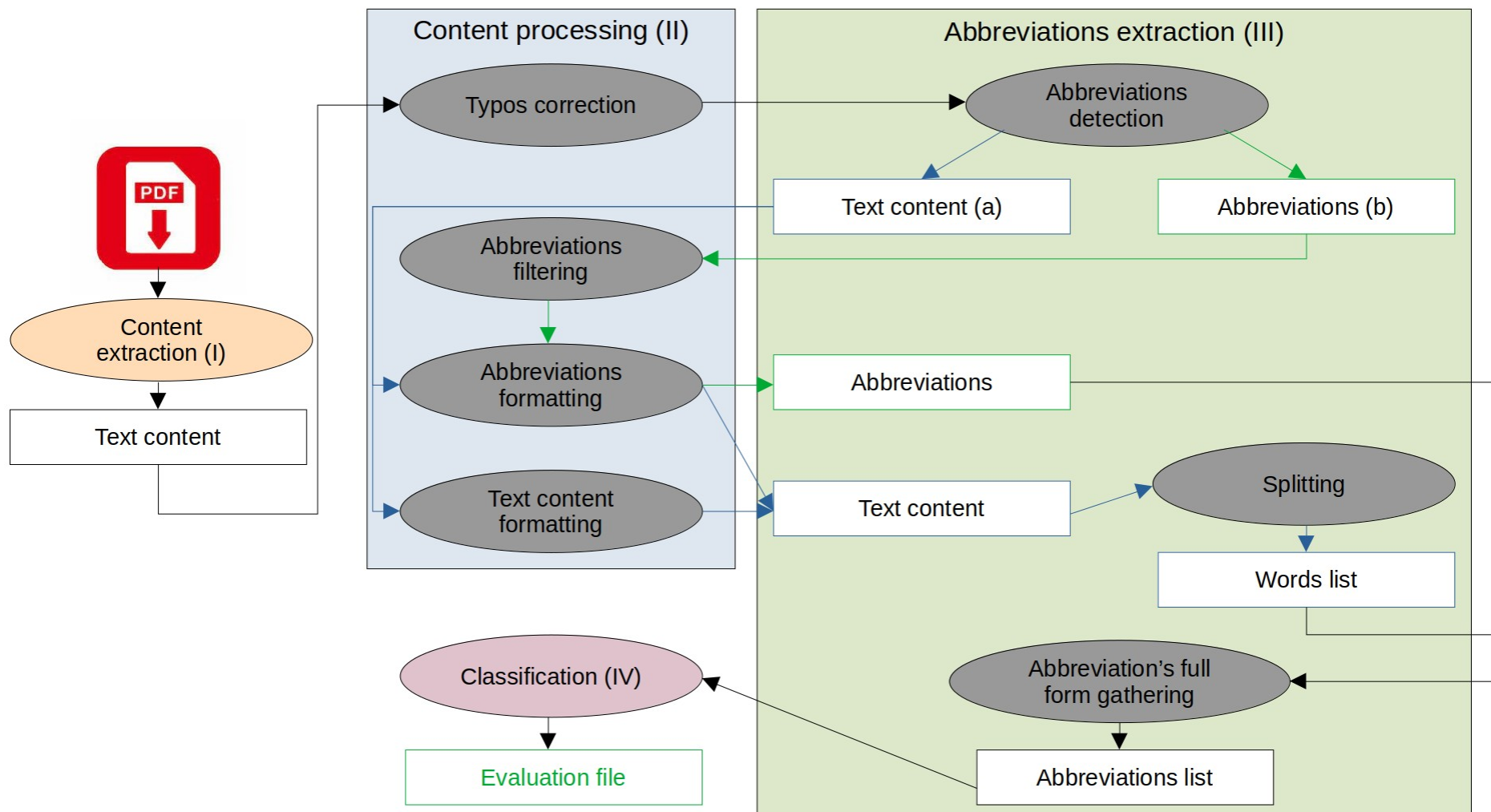
Between development and evaluation datasets



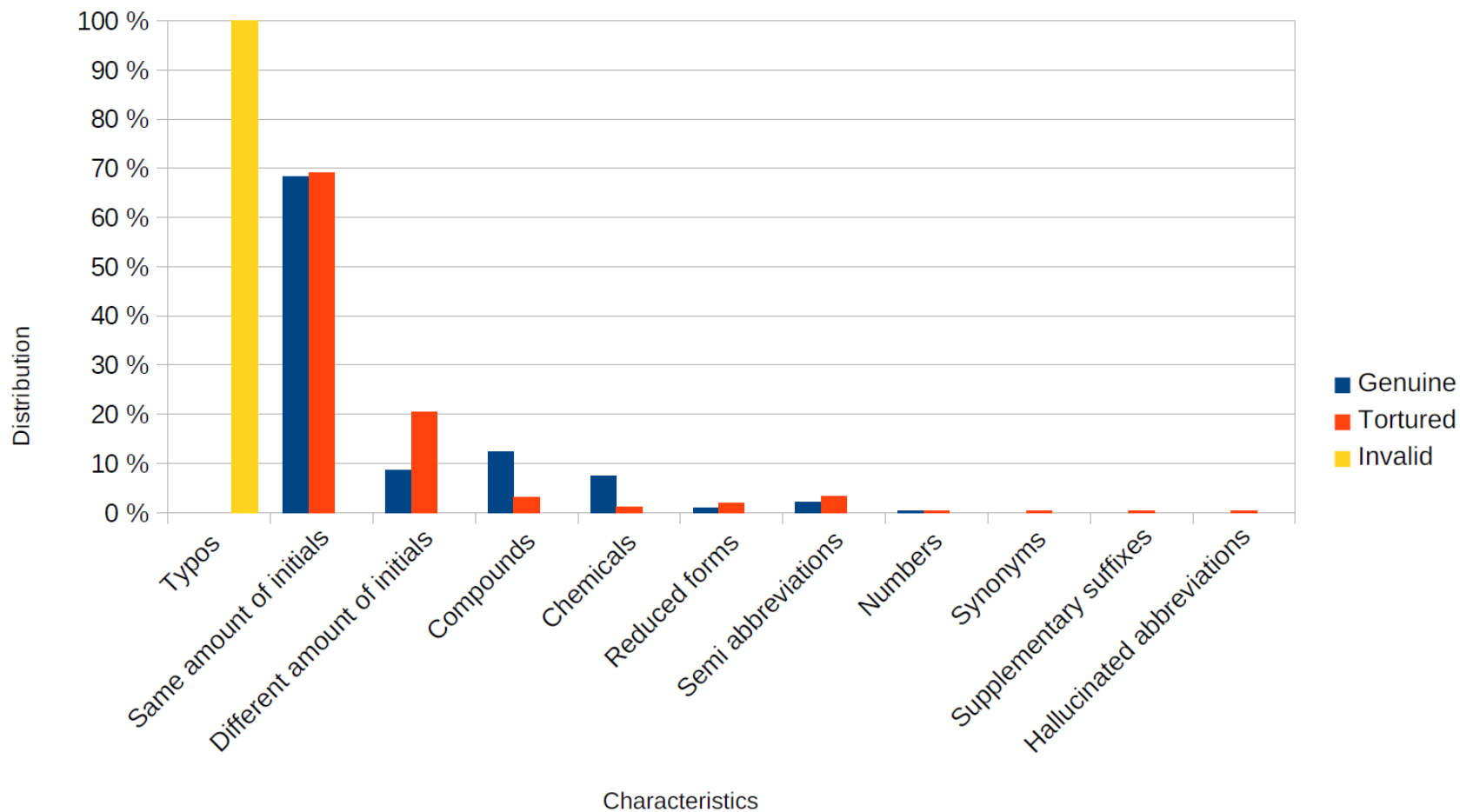
Between evaluation subsets



Extraction task



Abbreviations characteristics distribution



Characteristic	Class	
	Genuine	Tortured
(1) Same amount of initials	<i>Hidden Markov Model (HMM)</i>	<i>Concealed Markov Display (HMM)</i>
(2) Different amount of initials	<i>Generalised Linear Model (GLM)</i>	<i>Summed Up Direct Models (GLM)</i>
(3) Compounds	<i>Multidrug Resistance (MDR)</i>	<i>Multidrug Safe (MDR)</i>
(3) Chemicals	<i>Diethylenetriamine Pentaacetic acid (DTPA)</i>	<i>Diethylenetriamine Pentaacetic ruinous (DTPA)</i>
(4) Reduced forms	<i>Residual neural Network (ResNet)</i>	<i>Lingering brain Organization (ResNet)</i>

Characteristic	Class
(1) Same amount of initials	// (1) no offset using <u>Hidden Markov Model</u> (HMM) using <u>Concealed Markov Display</u> (HMM) → Genuine → Next step
(2) Different amount of initials	
(3) Compounds	// (2) no offset <u>the Moroccan Agency of Press</u> (MAP) → Genuine <u>the Moroccan Organization of Press</u> (MAP) → Next step
(3) Chemicals	
(4) Reduced forms	// (1) 1 token offset <u>Hidden Markov Model</u> architecture (HMM) → Genuine <u>Concealed Markov Display</u> architecture (HMM) → Next step
	// (2) 1 token offset <u>Center for Disease Control and Prevention</u> (CDC) → Genuine <u>Center for Illness Control and Prevention</u> (CDC) → Next step
	// (3) <u>[...] Multidrug Resistance [...]</u> (MDR) → Genuine <u>[...] Multidrug Safe [...]</u> (MDR) → Next step
	// (4) <u>[...] Residual Network [...]</u> (ResNet) → Genuine <u>[...] Lingering Organization [...]</u> (ResNet) → Tortured

- **Recall (R)**: $TP / (TP + FN)$
- **Precision (P)**: $TP / (TP + FP)$
- **F-score (F1)**: $2 * (P * R) / (P + R)$
- **Invalid abbreviations are considered as genuine**

Abbreviations extraction		
	Annotation	Prediction
TP	Abbreviation	Abbreviation
FP	-	Abbreviation
TN	Abbreviation	-
FN	-	-

Abbreviations classification		
	Annotation	Prediction
TP	Tortured	Tortured
FP	Genuine	Tortured
TN	Genuine	Genuine
FN	Tortured	Genuine

Tortured abbreviations extraction		
	Annotation	Prediction
TP	Tortured	Tortured
FP	Genuine	Tortured
	Genuine	-
TN	-	Tortured
	Tortured	Genuine
FN	Tortured	-
	-	Genuine
FN	-	-

Dataset	Task								
	Abbreviations extraction			Abbreviations classification			Tortured abbreviations extraction		
	R	P	F1	R	P	F1	R	P	F1
PPS_75	0.98	0.66	0.79	0.96	0.93	0.94	0.66	0.53	0.59
H_75	0.82	0.83	0.82	0.91	0.49	0.63	0.61	0.27	0.38
H_G25	0.88	0.79	0.83	0.55	0.05	0.10	0.21	0.03	0.05
H_R25	0.79	0.88	0.83	0.93	0.71	0.81	0.68	0.45	0.54
H_T25	0.76	0.82	0.79	0.94	0.80	0.87	0.64	0.48	0.55

→ **Rework the filtering rules** (hallucinated and polysemic expressions)

- **Domain agnostic**
- **Scalable**
- **PDF-oriented**
- **No need for fingerprints**
- Found **185 new fingerprints** for the PPS
(as of 04.06.2024, using 28 randomly selected **Hindawi** articles)

The case of **IOP Publishing**:

- Publications between **2017** and **2023**
- **3848 problematic** publications
- **10%** of problematic publications are **tortured**
- At least **7%** contain **tortured abbreviations**

Source: the PPS (consulted on May 21, 2024)

Hallucinated abbreviation: **Bolster Vector Machine (BVM)** → **Support Vector Machine (SVM)**

Polysemic expression: **Profound Learning** (tortured in CS, genuine in psychology)

Perspectives



Draft

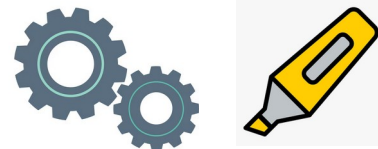


Preprint

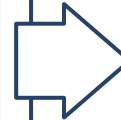


Article

Tortured Phrases Tool Kit (TPTK)



Tortured phrases extraction and highlighting



on IARPA's CS2 and NIST's IIR-An in both check and recognizable proof (for example search) and [9].

A half breed framework is introduced in which a convolutional neural organization and a logistic Regression Classifier are consolidated. Decreasing investigation is more effective when the ordinariness suspicious are identified. The thoughts are finished on Yale face information base shows improved arrangement rates in more modest measure of time [10]. Class Activated Mapping Gate-CAM is utilized to show the areas that our organization uses to segregate between [11].

Dense SIFT and customary SIFT are considered in [12] and contrasted when combined and CNN highlights. Besides, an aggregator of the models is created. Results show the prevalence of CNN with Dense SIFT over ordinary CNN and CNN with SIFT. The exactness even expanded when all the models are collected which produces confusion matrix results on 73.4% and 99.1% on FER-2013 and CK respectively.

We propose a grouping technique, called the **closest component line (NFL)** for face acknowledgment. Any two component purposes of individual face information by the element line (FL) going through the two focuses. The inferred FL catches big number of faces or pictures than the first focuses and in this manner extends the limit of the accessible information base. The characterization depends on the closest separation from the question includes highlight every FL. With a consolidated face information base, the NFL mistake rate is about 43.7-63.4% of that of the standard eigenface technique. Besides, the NFL achieves the most reduced mistake rate and answered to date for the ORL face information base [13].

The current works fail to announce the CNN designs that function admirably for face acknowledgment instead of examine the explanation. In this network, we address this issue by presenting CNN-based face acknowledgment framework (CNN-FRS) on a shared belief to make our work effectively reproducible. The three principal structures of CNN designs were utilizing LFW information. This paper quantitatively thinks about the designs of CNNs and assesses the impact of various usage decisions. We recognize a few helpful hypotheses of CNN-FRS. For example, the dimensionality of the **scholarly highlights** can be altogether decreased without unfriendly impact on face acknowledgment exactness. Likewise, a conventional measurement learning technique showing CNN-learned highlights is assessed. Analyses show two critical components to great CNN-FRS execution are the combination of numerous CNNs and metric learning [14].

A convolution neural organization (CNN) with **consideration instrument** that can see the impediment districts of the face and spotlight on the most discriminative non-blocked locales. Thinking about various RoIs, we present two realizations of ACNN. ACNN just focuses on neighborhood facial patches. ACNN incorporates nearby portrayals at fix level with worldwide portrayal at picture level. Trial results show that ACNNs improve the acknowledgment precision on both themon-impediment and multiblock datasets. Perception results show that, contrasted and the CNN without Gate Unit, ACNNs are fit for moving the consideration from the blocked patches to other related yet unimpeded ones. ACNNs additionally outperform other best-in-class techniques on face and video in the labourer appearance datasets under the cross-dataset assessment convention [15].

With the quickly advancement of the face acknowledgment (FR), warm face acknowledgment has gotten expanding consideration. In any case, the conventional techniques for warm face acknowledgment for the most part focus on the hand-made component plan, which requires more endeavour. For example, face feature extraction is a time-consuming task

has moderately lower acknowledgment rate. A convolutional neural organization (CNN) design for warm face acknowledgment is presented. CNN is another kind of neural organization strategy which can consequently take in viable highlights from the study information. Investigation results on FER-2013 face information base show that our proposed CNN engineering accomplishes higher acknowledgment rate contrasted and the customary acknowledgment, for example, LBP, HOG and minutiae invariant [16].

Video-based feeling acknowledgment framework joins repetitive neural organization (RNN) and 3D convolutional networks (3D) in a late-combination style. RNN and 3D encode appearance and movement data in an unexpected way. Particular RNN takes appearance highlights removed by convolutional neural organization (CNN) over individual video outlines as info and encodes movement later, while 3D models appearance and movement of video at the same time. Joined with a sound module, our framework accomplished an acknowledgment exactness of 59.02% without utilizing any extra feeling marked video cuts preparing set, contrasted with 53.8% of the state-of-the-art [17].

Brood examinations show that consolidating RNN and 3D together can improve video-based feeling acknowledgment perceptibility [17].

Perceiving human appearances is quite possibly the most well-known issues in the field of computer vision. Numerous methodologies and strategies have been tried and applied on the theme, particularly neural organizations. This paper proposed another method that can be supplemented at the lower part of a neural organization engineering as far as face acknowledgment, called the **inattention layer (ILL)**. To make more sure forecasts and arrangements, this inattention layer helps the profound learning model to indicate further discernible groups between various individuals (classes) by setting additional requirements on pictures of **face** similar individual (item-individual) while skipping pictures on pictures of **background** (between individual). This proposed obliged the inattention layer improved the acknowledgment exactness on countenances by 2% [19].

Face acknowledgment is critical to certifiable applications, for example, video observation, human machine connection and security frameworks. When contrasted with conventional AI (draw, net), profound learning based techniques have indicated better exhibitions regarding precision and speed of preparing in picture acknowledgment. This paper proposes an altered Convolutional Neural Network (CNN) design by adding two standardization activities to two of the layers. The standardization activity which is bunch standardization gave increasing speed of the organization. CNN design was utilized to extract particular face highlights and Softmax classifier was utilized to characterize faces in the **completely associated layer** of CNN. In the trial part, Georgia Tech Database indicated that the proposed approach has improved the **face acknowledgment exactness** with better acknowledgment results [21].

Face pictures showing up in sight and sound applications such as secure payment systems through cloud [21] [22], informal organizations and computerized management ordinarily display emotional posture, brightening, and appearance varieties, bringing about significant execution debasement for conventional face acknowledgment calculations. This paper proposes a complete **profound learning** structure to get the face appearance variations in real world data. The proposed **profound learning** structure is made out of a bunch of extravagantly planned convolutional neural organizations (CNNs) and a three-layer stacked auto-encoder (SAE). The arrangement of CNNs separates consecutive **facial highlights** from multi-dimensional information. At that point, these extracted highlights are connected to frame a high-dimensional component vector, whose measurements are compacted by SAE. The entropy of the CNNs are prepared utilizing a subset of 2000 subject, from the face and expression CASIA-WebFace information base, which are generated by a professional key

<https://pubpeer.com/publications/2432C53B542C6B29AF449894368D61>

Submission (authors)

Editorial workflow (publishers)

Decision making (editors)

- [1] Barbour B., Stell B. M., “PubPeer: Scientific Assessment Without Metrics”, in M. Biagioli, A. Lippman (ed.), *Gaming the Metrics: Misconduct and Manipulation in Academic Research*, The MIT Press, p. 149–155, 2020.
DOI : <https://doi.org/10.7551/mitpress/11087.003.0015>.
- [2] Becker J., Wahle J. P., Ruas T., Gipp B., “Paraphrase Detection: Human vs. Machine Content”, *Eighth Annual Conference on Advances in Cognitive Systems*, *Advances in Cognitive Systems*, 2020.
URL : https://advancesincognitivesystems.github.io/acs/data/ACS2020_paper_28.pdf.
- [3] Biagioli M., Lippman A. (ed.), *Gaming the Metrics: Misconduct and Manipulation in Academic Research*, The MIT Press, 2020.
DOI : <https://doi.org/10.7551/mitpress/11087.001.0001>.
- [4] Cabanac G., Labbe C., Magazinov A., “Tortured phrases: A dubious writing style emerging in science. Evidence of critical issues affecting established journals”, 2021. Preprint arXiv.
DOI : <https://doi.org/10.48550/arXiv.2107.06751>.
- [5] Cabanac G., Labbe C., Magazinov A., “The ‘Problematic Paper Screener’ automatically selects suspect publications for post-publication (re)assessment”, 2022. Preprint arXiv.
DOI : <https://doi.org/10.48550/arXiv.2210.04895>.
- [6] Clausse A., Cabanac G., Cuxac P., Labbe C., “Extraction d’acronymes tortures dans la littérature scientifique”, in P. Cuxac, C. Lopez (ed.), *Atelier TextMine de la conférence Extraction et Gestion des Connaissances (EGC) de 2024*, Dijon, France, p. 27–37, 2024.
URL : <https://hal.science/hal-04426448>.
- [7] Gehrman S., Strobelt H., Rush A., “GLTR: Statistical Detection and Visualization of Generated Text”, in M. R. Costa-jussa, E. Alfonseca (ed.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, Association for Computational Linguistics, p. 111–116, 2019.
DOI : <https://doi.org/10.18653/v1/P19-3019>.
- [8] Lay P., Lentschat M., Labbe C., “Investigating the detection of Tortured Phrases in Scientific Literature”, in A. Cohan, G. Feigenblat, D. Freitag, T. Ghosal, D. Herrmannova, P. Knoth, K. Lo, P. Mayr, M. Shmueli-Scheuer, A. de Waard, L. L. Wang (ed.), *Proceedings of the Third Workshop on Scholarly Document Processing*, Association for Computational Linguistics, p. 32–36, 2022.
URL : <https://aclanthology.org/2022.sdp-1.4>.
- [9] Macbeth J. C., Chang E., Chen J. G., Grandic S., “A Broader Range for ‘Meaning the Same Thing’: Human Against Machine on Hard Paraphrase Detection Tasks”, 2020.
URL : <https://api.semanticscholar.org/CorpusID:232247625>.
- [10] Martel E., Lentschat M., Labbe C., “Detection of Tortured Phrases in Scientific Literature”, in T. Ghosal, F. Grezes, T. Allen, K. Lockhart, A. Accomazzi, S. Blanco-Cuaresma (ed.), *Proceedings of the Second Workshop on Information Extraction from Scientific Publications*, Association for Computational Linguistics, p. 43–48, 2023.
DOI : <https://doi.org/10.18653/v1/2023.wiesp-1.6>.
- [11] Wahle J. P., Ruas T., Foltyněk T., Meuschke N., Gipp B., “Identifying Machine-Paraphrased Plagiarism”, in M. Smits (ed.), *Information for a Better World: Shaping the Global Future*, Springer, p. 393–413, 2022.
DOI : https://doi.org/10.1007/978-3-030-96957-8_34.

Text length before: 775

775/10,000

In this technique, the learning process is based on semi-labeled datasets. Occasionally, generative adversarial networks (GANs) and DRL are employed in the same way as this technique. In addition, RNNs, which include GRUs and LSTMs, are also employed for partially supervised learning. One of the advantages of this technique is to minimize the amount of labeled data needed. On the other hand, One of the disadvantages of this technique is irrelevant input feature present training data could furnish incorrect decisions. Text document classifier is one of the most popular example of an application of semi-supervised learning. Due to difficulty of obtaining a large amount of labeled text documents, semi-supervised learning is ideal for text document classification task.

Text length after: 756



Basic Spin

In this strategy, the educational experience depends on semi-named datasets. Sometimes generative ill-disposed networks (GANs) and DRL are utilized similarly as this procedure. What's more, RNNs, which incorporate GRUs and LSTMs, are additionally utilized for somewhat managed learning. One of the upsides of this procedure is to limit how much named information required. On the other hand, One of the drawbacks of this procedure is insignificant info include present preparation information could outfit erroneous choices. Text report classifier is one of the most famous illustration of a utilization of semi-administered learning. Because of trouble of getting a lot of marked text reports, semi-regulated learning is great for text record order task.



What is a paper mill?



Paper mills, as the name suggests, are organizations that profit by producing and selling fake research papers to desperate scientists who are likely under pressure to publish. Lately, they even sell authorship and also target scholarly journals editors to ensure paper acceptance.

#R3ReadToday# adapted from F. Joelsing, 2024
Paper mills are bribing editors at scholarly journals

Annex 3: PPPR using the PPS and PubPeer

- ▼ Detectors:
- 0 Annulled
- 1 Tortured**
- 2 SCiGen
- 3 Mathgen
- 4 SBIR
- 5 Suspect
- 6 Seek&Blastn
- 7 Problematic Cell Lines
- 8 Citejacked
- 9 Feet of Clay

1

Year	Doctype	Publisher	Venue	Title	Citations	Tips	Tortured Phrases	Assessor	PubPeer Users
2023	article	Wiley	<i>Computational Intelligence and Neuroscience</i>	Lung Cancer Classification in Histopathology Images Using Multiresolution Efficient Nets	3	11	bosom AND breast • component guide AND CNN • malignant growth AND cancer • bosom disease • completely associated layer • convolucional neural organization • equal calculation • forecast exactness • fully associated layer • neural organization • normal	Cabanac	Alexandre Clausse

2

#1 Alexandre Clausse comment accepted March 2024

3



This Hindawi article contains several **tortured phrases** that make some passages hard to parse. These typically result from an attempt to avoid plagiarism detection using a paraphrasing software. So far, the following have been spotted:

Tortured Phrases (found)	Established Phrases (expected)
bosom disease	breast cancer
bosom AND breast	breast
completely associated layer	fully connected layer (neural network)
component guide AND CNN	feature map
convolucional neural organization	convolucional neural network
equal calculation	parallel computing
forecast exactness	prediction accuracy
fully associated layer	fully connected layer (in a neural network)
malignant growth AND cancer	cancer
neural organization	neural network
cerebrum AND bosom	brain (tumor) AND breast (cancer)

Can the authors explain why they departed from the established phrases?

Sometimes the Colab is not assigning GPU to our notebook or cannot connect with the host, so we run it without GPU, and it takes many hours to train the model. Convolutional neural organizations (CNN) are neural organizations that are especially fit for picture classification. It has been, for instance, effectively utilized for picture classification orders [25–27]. A typical CNN design contains convolutional, pooling, and completely associated layers. Moderately novel procedures, for example, batch standardization [23],

How come these incorrect wordings survived proofreading by the coauthors, editors, referees, copy editors, and typesetters?

Flagged by the *Problematic Paper Screener*.

Characteristics	Class		
	Genuine	Tortured	Invalid
Typos	-	-	Tru_-Positive (TP)
Same amount of initials	<i>Pain Dysfunction Syndrome (PDS)</i>	<i>Concealed Markov Display (HMM)</i>	-
Different amount of initials	<i>Centers for Diseases Control and Prevention (CDC)</i>	<i>Summed Up Direct Models (GLM)</i>	-
Compounds	<i>Intraocular Pressure (IOP)</i>	<i>Multidrug Safe (MDR)</i>	-
Chemicals	<i>3-Aminopropyltriethoxysilane (3-APTES)</i>	<i>Diethylenetriamine Pentaacetic ruinous (DTPA)</i>	-
Reduced forms	<i>Residual neural Networks (ResNets)</i>	<i>Lingering brain Organization (ResNet)</i>	-
Semi-abbreviations	<i>Gabor-HOG (GHOG)</i>	<i>Non-straight ARMA models (NARMA)</i>	-
Numbers	<i>Three-Dimensional (3D)</i>	<i>Fifth-Age (5G)</i>	-
Synonyms	-	<i>Human PC Interface/Cooperation</i>	-
Supplementary suffixes	-	<i>PC Assisted Determination (CADx)</i>	-
Hallucinated abbreviations	-	<i>Bolster Vector Machine (BVM)</i>	-